Characterization of additive gene–environment interactions for colorectal cancer risk

Epidemiology

Thomas, Claire E.; Lin, Yi; Kim, Michelle; Kawaguchi, Eric S.; Qu, Conghui et al

https://doi.org/10.1097/EDE.0000000000001795

Characterization of Additive Gene–environment Interactions For Colorectal Cancer Risk

Claire E. Thomas,[a] Yi Lin,[a] Michelle Kim,[a] Eric S. Kawaguchi,[b] Conghui Qu,[a] Caroline Y. Um,[c] Brigid M. Lynch,[d,e] Bethany Van Guelpen,[f,g] Kostas Tsilidis,[h,i] Robert Carreras-Torres,[j,k] Franzel J.B. van Duijnhoven,[l] Lori C. Sakoda,[a,m] Peter T. Campbell,[n] Yu Tian,[o,p] Jenny Chang-Claude,[q,r] Stéphane Bézieau,[s] Arif Budiarto,[t,u] Julie R. Palmer,[v] Polly A. Newcomb,[a,w] Graham Casey,[x] Loic Le Marchandz,[y] Marios Giannakis,[z,aa,ab] Christopher I. Li,[a] Andrea Gsur,[ac] Christina Newton,[c] Mireia Obón-Santacana,[ad,ae,af] Victor Moreno,[ad,ae,af,ag] Pavel Vodicka,[ah,ai,aj] Hermann Brenner,[ak,al,am] Michael Hoffmeister,[ak] Andrew J. Pellatt,[an] Robert E. Schoen,[ao] Niki Dimou,[ap] Neil Murphy,[ap] Marc J. Gunter,[ap] Sergi Castellví-Bel,[aq] Jane C. Figueiredo,[ar,as] Andrew T. Chan,[at,au,av,aw,ax,ay] Mingyang Song,[av,az,aaa] Li Li,[aab] D. Timothy Bishop,[aac] Stephen B. Gruber,[aad] James W. Baurley,[aae,aaf] Stephanie A. Bien,[a] David V. Conti,[b] Jeroen R. Huyghe,[a] Anshul Kundaje,[aag,aah] Yu-Ru Su,[a] Jun Wang,[aai,aaj] Temitope O. Keku,[aak] Michael O. Woods,[aal] Sonja I. Berndt,[aam] Stephen J. Chanock,[aam] Catherine M. Tangen,[aan] Alicja Wolk,[aao] Andrea Burnett-Hartman,[aap] Anna H. Wu,[aaq] Emily White,[a,aar] Matthew A. Devall,[aas] Virginia Díez-Obrero,[aat] David A Drew,[aau] Edward Giovannucci,[aav,aaw] Akihisa Hidaka,[a] Andre E. Kim,[b] Juan Pablo Lewinger,[b] John Morrison,[b] Jennifer Ose,[aax,aay] Nikos Papadimitriou,[ap] Bens Pardamean,[aae] Anita R. Peoples,[aax,aay] Edward A. Ruiz-Narvaez,[aaz] Anna Shcherbina,[aag,aah] Mariana C. Stern,[as] Xuechen Chen,[ak] Duncan C. Thomas,[b] Elizabeth A. Platz,[aaaa] W. James Gauderman,[b] Ulrike Peters,[a,w] and Li Hsu[a,aaab]

Submitted December 4, 2023; accepted September 18, 2024

From the [a]Public Health Sciences Division, Fred Hutchinson Cancer Center, Seattle, WA; [b]Division of Biostatistics, Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA; [c]Department of Population Science, American Cancer Society, Atlanta, GA; [d]Cancer Epidemiology Division, Cancer Council Victoria, Melbourne, Victoria, Australia; [e]Centre for Epidemiology and Biostatistics, Melbourne School of Population and Global Health, The University of Melbourne, Victoria, Australia; [f]Department of Radiation Sciences, Oncology Unit, Umeå University, Umeå, Sweden; [g]Wallenberg Centre for Molecular Medicine, Umeå University, Umeå, Sweden; [h]Department of Epidemiology and Biostatistics, Imperial College London, School of Public Health, London, UK; [i]Department of Hygiene and Epidemiology, University of Ioannina, School of Medicine, Ioannina, Greece; [j]Colorectal Cancer Group, ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain; [k]Digestive Diseases and Microbiota Group, Girona Biomedical Research Institute (IDIBGI), Salt, Girona, Spain; [l]Division of Human Nutrition and Health, Wageningen University & Research, Wageningen, The Netherlands; [m]Division of Research, Kaiser Permanente Northern California, Oakland, CA; [n]Department of Epidemiology and Population Health, Albert Einstein College of Medicine, Bronx, NY; [o]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; [p]School of Public Health, Capital Medical University, Beijing, China; [q]Division of Cancer Epidemiology, German Cancer Research Center (DKFZ), Heidelberg, Germany; [r]University Medical Centre Hamburg-Eppendorf, University Cancer Centre Hamburg (UCCH), Hamburg, Germany; [s]Service de Génétique Médicale, Centre Hospitalier Universitaire (CHU) Nantes, Nantes, France; [t]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia; [u]Computer Science Department, School of Computer Science, Bina Nusantara University, Jakarta, Indonesia; [v]Slone Epidemiology Center at Boston University, Boston, MA; [w]Department of Epidemiology, University of Washington, Seattle, WA; [x]Center for Public Health Genomics, University of Virginia, Charlottesville, VA; [y]University of Hawaii Cancer Center, Honolulu, HI; [z]Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA; [aa]Broad Institute of MIT and Harvard, Cambridge, MA; [ab]Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA; [ac]Center for Cancer Research, Medical University of Vienna, Vienna, Austria; [ad]Unit of Biomarkers and Suceptibility (UBS), Oncology Data Analytics Program (ODAP), Catalan Institute of Oncology (ICO), L'Hospitalet del Llobregat, Barcelona, Spain; [ae]ONCOBELL Program, Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat, Barcelona, Spain; [af]Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), Madrid, Spain; [ag]Department of Clinical Sciences, Faculty of Medicine and health Sciences and Universitat de Barcelona Institute of Complex Systems (UBICS), University of Barcelona (UB), L'Hospitalet de Llobregat, Barcelona, Spain; [ah]Department of Molecular Biology of Cancer, Institute of Experimental Medicine of the Czech Academy of Sciences, Prague, Czech Republic; [ai]Institute of Biology and Medical Genetics, First Faculty of Medicine, Charles University, Prague, Czech Republic; [aj]Faculty of Medicine and Biomedical Center in Pilsen, Charles University, Pilsen, Czech Republic; [ak]Division of Clinical Epidemiology and Aging Research, German Cancer Research Center (DKFZ), Heidelberg, Germany; [al]Division of Preventive Oncology, German Cancer Research Center (DKFZ) and National Center for Tumor Diseases (NCT), Heidelberg, Germany; [am]German Cancer Consortium (DKTK), German Cancer Research Center (DKFZ), Heidelberg, Germany; [an]Department of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, TX; [ao]Departments of Medicine and Epidemiology, University of Pittsburgh Medical Center, Pittsburgh, PA; [ap]Nutrition and Metabolism Branch, International Agency for Research on Cancer, Lyon, France; [aq]Gastroenterology Department, Hospital Clínic, Institut d'Investigacions Biomèdiques

ISSN: 1044-3983/25/361-126138
DOI: 10.1097/EDE.0000000000001795

August Pi i Sunyer (IDIBAPS), Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas (CIBEREHD), University of Barcelona, Barcelona, Spain; [ar]Department of Medicine, Samuel Oschin Comprehensive Cancer Institute, Cedars-Sinai Medical Center, Los Angeles, CA; [as]Department of Population and Public Health Sciences, Keck School of Medicine, University of Southern California, Los Angeles, CA; [at]Division of Gastroenterology, Massachusetts General Hospital and Harvard Medical School, Boston, MA; [au]Channing Division of Network Medicine, Brigham and Women's Hospital and Harvard Medical School, Boston, MA; [av]Clinical and Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA; [aw]Broad Institute of Harvard and MIT, Cambridge, MA; [ax]Department of Epidemiology, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA; [ay]Department of Immunology and Infectious Diseases, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA; [az]Departments of Epidemiology and Nutrition, Harvard TH Chan School of Public Health, Boston, MA; [aaa]Department of Nutrition, Harvard T.H. Chan School of Public Health, Harvard University, Boston, MA; [aab]Department of Family Medicine, University of Virginia, Charlottesville, VA; [aac]Leeds Institute of Medical Research, University of Leeds, Leeds, UK; [aad]Department of Medical Oncology & Therapeutics Research, City of Hope National Medical Center, Duarte CA; [aae]Bioinformatics and Data Science Research Center, Bina Nusantara University, Jakarta, Indonesia; [aaf]BioRealm LLC, Walnut, CA; [aag]Department of Genetics, Stanford University, Stanford, CA; [aah]Department of Computer Science, Stanford University, Stanford, CA; [aai]Department of Preventive Medicine, Keck School of Medicine, University of Southern California, Los Angeles, CA; [aaj]Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA; [aak]Center for Gastrointestinal Biology and Disease, University of North Carolina, Chapel Hill, NC; [aal]Memorial University of Newfoundland, Discipline of Genetics, St. John's, Canada; [aam]Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Bethesda, MD; [aan]SWOG Statistical Center, Fred Hutchinson Cancer Center, Seattle, WA; [aao]Institute of Environmental Medicine, Karolinska Institutet, Stockholm, Sweden; [aap]Institute for Health Research, Kaiser Permanente Colorado, Aurora, CO; [aaq]University of Southern California, Department of Population and Public Health Sciences, Los Angeles, CA; [aar]Department of Epidemiology, University of Washington School of Public Health, Seattle, WA; [aas]Department of Family Medicine, University of Virginia, Charlottesville, VA; [aat]Novo Nordisk Foundation Center for Basic Metabolic Research, Faculty of Health and Medical Sciences, University of Copenhagen, Denmark; [aau]Clinical & Translational Epidemiology Unit, Massachusetts General Hospital and Harvard Medical School, Boston, MA; [aav]Harvard TH Chan School of Public Health; [aaw]Department of Biostatistics, University of Washington, Seattle, WA; [aax]Huntsman Cancer Institute, Salt Lake City, UT; [aay]Department of Population Health Sciences, University of Utah, Salt Lake City, UT; [aaz]Department of Nutritional Sciences, University of Michigan School of Public Health, Ann Arbor, MI; [aaaa]Department of Epidemiology, Johns Hopkins Bloomberg School of Public Health, Baltimore, MD; and [aaab]Department of Biostatistics, University of Washington, Seattle, WA.

Genetics and Epidemiology of Colorectal Cancer Consortium (GECCO): National Cancer Institute, National Institutes of Health, U.S. Department of Health and Human Services (U01 CA137088, R01 CA059045, U01 CA164930, R21 CA191312, R01 CA244588, R01 206279, R01 201407, R01 CA488857, P20 CA252733). Genotyping/Sequencing services were provided by the Center for Inherited Disease Research (CIDR) contract number HHSN2682017000006I and HHSN2682012000008I. This research was funded in part through the NIH/NCI Cancer Center Support Grant P30 CA015704. Scientific Computing Infrastructure at Fred Hutch funded by ORIP grant S10OD028685. C.E.T. is supported by L70 CA284301, T32 CA094880, and T32 CA009168. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Cancer Institutes or the National Institutes of Health.

ASTERISK: a Hospital Clinical Research Program (PHRC-BRD09/C) from the University Hospital Center of Nantes (CHU de Nantes) and supported by the Regional Council of Pays de la Loire, the Groupement des Entreprises Françaises dans la Lutte contre le Cancer (GEFLUC), the Association Anne de Bretagne Génétique and the Ligue Régionale Contre le Cancer (LRCC).

The ATBC Study is supported by the Intramural Research Program of the US National Cancer Institute, National Institutes of Health, Department of Health and Human Services.

BWHS: National Institutes of Health U01 CA164974.

The Colon Cancer Family Registry (CCFR, www.coloncfr.org) is supported in part by funding from the National Cancer Institute (NCI), and National Institutes of Health (NIH) (award U01 CA167551). Support for case ascertainment was provided in part by the Surveillance, Epidemiology, and End Results (SEER) Program and the following US state cancer registries: AZ, CO, MN, NC, NH; and by the Victoria Cancer Registry (Australia) and Ontario Cancer Registry (Canada). The CCFR Set-1 (Illumina 1M/1M-Duo) was supported by NIH awards U01 CA122839 and R01 CA143237 (to G.C.). The CCFR Set-3 (Affymetrix Axiom CORECT Set array) was supported by NIH award U19 CA148107 and R01 CA81488 (to S.G.). The CCFR Set-4 (Illumina OncoArray 600K SNP a.rray) was supported by NIH award U19 CA148107 (to S.G.) and by the Center for Inherited Disease Research (CIDR), which is funded by the NIH to the Johns Hopkins University, contract number HHSN2682012000008I. The content of this manuscript does not necessarily reflect the views or policies of the NCI, NIH, or any of the collaborating centers in the Colon Cancer Family Registry (CCFR), nor does mention of trade names, commercial products, or organizations imply endorsement by the US Government, any cancer registry, or the CCFR.

CLUE II funding was from the National Cancer Institute (U01 CA086308, Early Detection Research Network; P30 CA006973), National Institute on Aging (U01 AG018033), and the American Institute for Cancer Research. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government.

COLO2&3: National Institutes of Health (R01 CA060987).

ColoCare: This work was supported by the National Institutes of Health (grant numbers R01 CA189184 (Li/Ulrich), U01 CA206110 (Ulrich/Li/Siegel/Figueiredo/Colditz, 2P30CA015704- 40 (Gilliland), R01 CA207371 (Ulrich/Li)), the Matthias Lackas-Foundation, the German Consortium for Translational Cancer Research, and the EU TRANSCAN initiative.

CORSA: The CORSA study was funded by Austrian Research Funding Agency (FFG) BRIDGE (grant 829675, to Andrea Gsur), the "Herzfelder'sche Familienstiftung" (grant to Andrea Gsur) and was supported by COST Action BM1206.

CPS-II: The American Cancer Society funds the creation, maintenance, and updating of the Cancer Prevention Study-II (CPS-II) cohort. The study protocol was approved by the institutional review boards of Emory University, and those of participating registries as required.

CRCGEN: Colorectal Cancer Genetics & Genomics, Spanish study was supported by Instituto de Salud Carlos III, co-funded by FEDER funds—a way to build Europe—(grants PI14-613 and PI09-1286), Agency for Management of University and Research Grants (AGAUR) of the Catalan Government (grant 2017SGR723), Junta de Castilla y León (grant LE22A10-2), the Spanish Association Against Cancer (AECC) Scientific Foundation grant GCTRA18022MORE and the Consortium for Biomedical Research in Epidemiology and Public Health (CIBERESP), action Genrisk. Sample collection of this work was supported by the Xarxa de Bancs de Tumors de Catalunya sponsored by Pla Director d'Oncología de Catalunya (XBTC), Plataforma Biobancos PT13/0010/0013 and ICOBIOBANC, sponsored by the Catalan Institute of Oncology. We thank CERCA Programme, Generalitat de Catalunya for institutional support.

Czech Republic CCS: This work was supported by the Czech Science Foundation (21-04607X, 21-27902S), by the Grant Agency of the Ministry of Health of the Czech Republic (grants AZV NU21-07-00247 and AZV NU21-03-00145), and Charles University Research Fund (Cooperation 43-Surgical disciplines).

DACHS: This work was supported by the German Research Council (BR 1704/6-1, BR 1704/6-3, BR 1704/6-4, CH 117/1-1, HO 5117/2-1, HE 5998/2-1, KL 2354/3-1, RO 2270/8-1, and BR 1704/17-1), the Interdisciplinary Research Program of the National Center for Tumor Diseases (NCT), Germany, and the German Federal Ministry of Education and Research (01KH0404, 01ER0814, 01ER0815, 01ER1505A, and 01ER1505B).

DALS: National Institutes of Health (R01 CA048998 to M. L. Slattery).

EDRN: This work is funded and supported by the NCI, EDRN Grant (U01-CA152753).

EPIC: The coordination of EPIC is financially supported by the International Agency for Research on Cancer (IARC) and also by the Department of Epidemiology and Biostatistics, School of Public Health, Imperial College London which has additional infrastructure support provided by the NIHR Imperial Biomedical Research Centre (BRC). The national cohorts are supported by: Danish Cancer Society (Denmark); Ligue Contre le Cancer, Institut Gustave Roussy, Mutuelle Générale de l'Education Nationale, Institut National de la Santé et de la Recherche Médicale (INSERM) (France); German Cancer Aid, German Cancer Research Center (DKFZ), German Institute of Human Nutrition Potsdam- Rehbruecke (DIfE), Federal Ministry of Education and Research (BMBF) (Germany); Associazione Italiana per la Ricerca sul Cancro-AIRC-Italy, Compagnia di SanPaolo and National Research Council (Italy); Dutch Ministry of Public Health, Welfare and Sports (VWS), Netherlands Cancer Registry (NKR), LK Research Funds, Dutch Prevention Funds, Dutch ZON (Zorg Onderzoek Nederland), World Cancer Research Fund (WCRF), Statistics Netherlands (The Netherlands); Health Research Fund (FIS)—Instituto de Salud Carlos III (ISCIII), Regional Governments of Andalucía, Asturias, Basque Country, Murcia and Navarra, and the Catalan Institute of Oncology—ICO (Spain); Swedish Cancer Society, Swedish Research Council and Region Skåne and Region Västerbotten (Sweden); Cancer Research UK (14136 to EPIC-Norfolk; C8221/A29017 to EPIC-Oxford), Medical Research Council (1000143 to EPIC-Norfolk; MR/M012190/1 to EPIC-Oxford). (United Kingdom).

EPICOLON: This work was supported by grants from Fondo de Investigación Sanitaria/FEDER (PI08/0024, PI08/1276, PS09/02368, P111/00219, PI11/00681, PI14/00173, PI14/00230, PI17/00509, 17/00878, PI20/00113, PI20/00226, Acción Transversal de Cáncer), Xunta de Galicia (PGIDIT07PXIB9101209PR), Ministerio de Economia y Competitividad (SAF07-64873, SAF 2010-19273, and SAF2014-54453R), Fundación Científica de la Asociación Española contra el Cáncer (GCB13131592CAST), Beca Grupo de Trabajo "Oncología" AEG (Asociación Española de Gastroenterología), Fundación Privada Olga Torres, FP7 CHIBCHA Consortium, Agència de Gestió d'Ajuts Universitaris i de Recerca (AGAUR, Generalitat de Catalunya, 2014SGR135, 2014SGR255, 2017SGR21, 2017SGR653), Catalan Tumour Bank Network (Pla Director d'Oncologia, Generalitat de Catalunya), PERIS (SLT002/16/00398, Generalitat de Catalunya), CERCA Programme (Generalitat de Catalunya) and COST Action BM1206 and CA17118. CIBERehd is funded by the Instituto de Salud Carlos III.

ESTHER/VERDI. This work was supported by grants from the Baden-Württemberg Ministry of Science, Research and Arts, and the German Cancer Aid.

Hawaii Adenoma Study: NCI grants R01 CA072520.

Harvard cohorts: HPFS is supported by the National Institutes of Health (P01 CA055075, UM1 CA167552, U01 CA167552, R01 CA137178, R01 CA151993, and R35 CA197735), NHS by the National Institutes of Health (P01 CA087969, UM1 CA186107, R01 CA137178, R01 CA151993, and R35 CA197735), and PHS by the National Institutes of Health (R01 CA042182).

Kentucky: This work was supported by the following grant support: Clinical Investigator Award from Damon Runyon Cancer Research Foundation (CI-8); NCI R01CA136726.

LCCS: The Leeds Colorectal Cancer Study was funded by the Food Standards Agency and Cancer Research UK Programme Award (C588/A19167).

MCCS cohort recruitment was funded by VicHealth and Cancer Council Victoria. The MCCS was further supported by Australian NHMRC grants 509348, 209057, 251553, and 504711 and by infrastructure provided by Cancer Council Victoria. Cases and their vital status were ascertained through the Victorian Cancer Registry (VCR) and the Australian Institute of Health and Welfare (AIHW), including the National Death Index and the Australian Cancer Database. BMLynch was supported by MCRF18005 from the Victorian Cancer Agency.

MEC: National Institutes of Health (R37 CA054281, P01 CA033619, and R01 CA063464).

MECC: This work was supported by the National Institutes of Health, U.S. Department of Health and Human Services (R01 CA081488, R01 CA197350, U19 CA148107, R01 CA242218, and a generous gift from Daniel and Maryann Fong.

NCCCS I & II: We acknowledge funding support for this project from the National Institutes of Health, R01 CA066635, and P30 DK034987.

NFCCR: This work was supported by an Interdisciplinary Health Research Team award from the Canadian Institutes of Health Research (CRT 43821); the National Institutes of Health, US Department of Health and Human Services (U01 CA074783); and the National Cancer Institute of Canada grants (18223 and 18226). The authors wish to acknowledge the contribution of Alexandre Belisle and the genotyping team of the McGill University and Génome Québec Innovation Centre, Montréal, Canada, for genotyping the Sequenom panel in the NFCCR samples. Funding was provided to M.W. by the Canadian Cancer Society Research Institute.

PLCO: This work was supported by the Intramural Research Program of the Division of Cancer Epidemiology and Genetics and by contracts from the Division of Cancer Prevention, National Cancer Institute, NIH, DHHS. Funding was provided by the National Institutes of Health (NIH), Genes, Environment and Health Initiative (GEI) Z01 CP 010200, NIH U01 HG004446, and NIH GEI U01 HG 004438.

REACH & SMS: This work was supported by the National Cancer Institute (grant P01 CA074184 to J.D.P. and P.A.N., grants R01 CA097325, R03 CA153323, and K05 CA152715 to P.A.N., and the National Center for Advancing Translational Sciences at the National Institutes of Health (grant KL2 TR000421 to A.N.B.-H.).

SEARCH: The University of Cambridge has received salary support in respect of PDPP from the NHS in the East of England through the Clinical Academic Reserve. Cancer Research UK (C490/A16561); the UK National Institute for Health Research Biomedical Research Centres at the University of Cambridge.

SELECT: Research reported in this publication was supported in part by the National Cancer Institute of the National Institutes of Health under Award Numbers U10 CA037429 (CD Blanke), and UM1 CA182883 (CM Tangen/IM Thompson). The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Swedish Mammography Cohort and Cohort of Swedish Men: This work is supported by the Swedish Research Council/Infrastructure grant, the Swedish Cancer Foundation, and the Karolinska Institute´s Distinguished Professor Award to Alicja Wolk.

UK Biobank: This research has been conducted using the UK Biobank Resource under Application Number 8614.

VITAL: National Institutes of Health (K05 CA154337).

WHI: The WHI program is funded by the National Heart, Lung, and Blood Institute, National Institutes of Health, U.S. Department of Health and Human Services through contracts HHSN268201100046C, HHSN268201100001C, HHSN268201100002C, HHSN268201100003C, HHSN268201100004C, and HHSN271201100004C.

S.G.: Brogent International LLC, co-founder, not related to submitted work. J.B.: co-founder and employee of BioRealm LLC. BioRealm LLC offers data analysis services, unrelated to this study. M.G.: Research funding from Servier and Janssen, unrelated to this study. Other authors report no conflicts of interest.

Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

Except for UK Biobank and CCFR 1 and CCFR 2, individual-level data are deposited in dbGaP (accession nos. phs001415.v1.p1, phs001315.v1.p1, phs001078.v1.p1, phs001903.v1.p1, phs001856.v1.p1, and phs001045.v1.p1). UK Biobank data are available through http://www.ukbiobank.ac.uk. CCFR 1 and CCFR 2 data can be requested by submitting an application for collaboration to the CCFR (forms, instructions, and contact information can be located at https://coloncfr.org/for-researchers/collaborate-with-the-ccfr/). Code is available from the corresponding author upon request.

SDC Supplemental digital content is available through direct URL citations in the HTML and PDF versions of this article (www.epidem.com).

Correspondence: Li Hsu and Ulrike Peters, Public Health Sciences Division, Fred Hutchinson Cancer Center, 1100 Fairview Ave N, Seattle, WA 98109. Email: lih@fredhutch.org; upeters@fredhutch.org.

Claire E Thomas and Yi Lin are considered as co-first author.

Background: Colorectal cancer (CRC) is a common, fatal cancer. Identifying subgroups who may benefit more from intervention is of critical public health importance. Previous studies have assessed multiplicative interaction between genetic risk scores and environmental factors, but few have assessed additive interaction, the relevant public health measure.

Methods: Using resources from CRC consortia, including 45,247 CRC cases and 52,671 controls, we assessed multiplicative and additive interaction (relative excess risk due to interaction, RERI) using logistic regression between 13 harmonized environmental factors and genetic risk score, including 141 variants associated with CRC risk.

Results: There was no evidence of multiplicative interaction between environmental factors and genetic risk score. There was additive interaction where, for individuals with high genetic susceptibility, either heavy drinking (RERI = 0.24, 95% confidence interval [CI] = 0.13, 0.36), ever smoking (0.11 [0.05, 0.16]), high body mass index (female 0.09 [0.05, 0.13], male 0.10 [0.05, 0.14]), or high red meat intake (highest versus lowest quartile 0.18 [0.09, 0.27]) was associated with excess CRC risk greater than that for individuals with average genetic susceptibility. Conversely, we estimate those with high genetic susceptibility may benefit more from reducing CRC risk with aspirin/nonsteroidal anti-inflammatory drugs use (−0.16 [−0.20, −0.11]) or higher intake of fruit, fiber, or calcium (highest quartile versus lowest quartile −0.12 [−0.18, −0.050]; −0.16 [−0.23, −0.09]; −0.11 [−0.18, −0.05], respectively) than those with average genetic susceptibility.

Conclusions: Additive interaction is important to assess for identifying subgroups who may benefit from intervention. The subgroups identified in this study may help inform precision CRC prevention.

Keywords: GxE; multiplicative interaction; additive interaction; colorectal cancer; genetic epidemiology

(Epidemiology 2025;36: 126–138)

Colorectal cancer (CRC) is a critical public health issue, as the third most commonly diagnosed cancer and the second leading cause of cancer death globally.[1] Ample evidence suggests that both genetics and environmental risk factors contribute to CRC development; heritability of CRC is estimated to be 15–35%.[2,3] Genome-wide association studies (GWAS) of CRC have identified important genetic risk associations, whereas recent GWAS studies of more than 100,000 participants have identified over 100 independent risk associations related to normal colorectal homeostasis, Hedgehog signaling, proliferation, cell adhesion, migration, immune function, long noncoding RNAs and somatic drivers, and microbial interactions.[4,5] There is great interest in identifying the interaction between modifiable exposures and genetic risk for precision prevention of CRC, pointing to environmental factors that could be targeted to mitigate elevated genetic risk.[6]

Gene–environment interaction (GxE) refers to when the effect of one exposure (environmental, E) on an outcome varies across different strata of the second exposure (genetic, G). For a disease trait, these interactions can be commonly described in two ways: additive and multiplicative. Additive interaction is focused on the sum of the individual effects of G and E, while multiplicative interaction is focused on the product of the individual effects. Given a logistic regression model, which is widely used to model GxE for the binary outcome from case–control studies, the departure from the multiplicative effect can be measured by the regression coefficient of the G and E product term ($\beta_{G\times E}$). The departure from the additive effect can be described with the relative excess risk due to interaction (RERI). This cannot be validly estimated directly from linear models with case–control data,[7] but it can be derived from estimates from the logistic regression model when the case–control odds ratio (OR) approximates the risk ratio. It has been asserted that interaction for the identification of relevant subgroups with excess risk and public health impact is best assessed through additive interaction, which is often overlooked in epidemiologic studies focused on etiology.[8] Multiplicative interaction is straightforward and convenient to assess through an interaction term in a logistic regression model that is easily derived in statistical software. It is important to note that when assessing interaction, if both exposures of interest have an effect on the outcome, then mathematically interaction must be present on at least one scale or both, where the absence of additive interaction implies multiplicative interaction, and vice versa.[8–10] Given that different information can be gained from different scales, it is recommended to present both additive and multiplicative interaction in practice.[8,10,11]

Many previous studies of GxE in CRC have assessed multiplicative interaction with genetic risk score,[12–19] but to our knowledge, only one study has included additive interaction using the RERI.[20] This study assessed interaction between a genetic risk score of 95 genetic variants and a healthy lifestyle score combining all environmental factors. However, it may have limited public health utility in identifying which component of a healthy lifestyle score particularly warrants intervention among those with high genetic risk. Therefore, we set out to assess GxE interaction among 97,918 participants of European descent utilizing a genetic risk score including 141 known CRC-associated genetic variants and 13 environmental risk/protective factors. Our objective was two-fold: to assess GxE interactions to identify subgroups among whom intervention on one or more environmental factors may have a large impact on reducing CRC risk, and to provide an example of GxE assessment on both the additive and multiplicative scales in agreement with current recommendations[8] to encourage the evaluation of additive interaction to guide precision prevention.

METHODS

Study Population

We included participants in the Colon Cancer Family Registry, the Colorectal Transdisciplinary Study, and the Genetics and Epidemiology of CRC Consortium (GECCO). Study details have been previously published[4,21,22] and can be found in eTable 1; http://links.lww.com/EDE/C195. Cases were

identified as incident invasive CRC or advanced adenoma cases and confirmed by medical records, pathology reports, or death certificate information. For cohort studies, we assembled nested case–control sets via risk-set sampling. For population-based case–control studies, we used population-based controls with study-specific eligibility. Controls were matched with cases on age, sex, race, and enrollment date or trial group, when applicable. For the subset of advanced adenoma cases (N = 4774, eTable 1; http://links.lww.com/EDE/C195), matched controls had polyp-free from endoscopy at the time of adenoma selection. All participants gave written informed consent, and studies were approved by their respective institutional review boards.

We limited analyses to self-reported European/non-Hispanic White participants who, as part of the quality control step for genetic association analyses, were also required to cluster with the European reference population from the 1000 Genomes Project based on genetic principal components. The final pooled sample size was 45,247 cases and 52,671 controls.

Environmental Factor Selection

Demographic, lifestyle, body composition, and environmental exposures, throughout the article, referred to as environmental factors, were self-reported either at in-person or telephone interviews or via structured questionnaires. For cohort studies, the variables were assessed at blood collection or participant recruitment; for case–control studies, variables were assessed at least 1–2 years and up to 10 years or more preceding participant recruitment. Data harmonization consisted of a multi-step procedure performed at the GECCO coordinating center (Fred Hutchinson Cancer Center).[23] Briefly, we defined common data elements a priori. We examined study questionnaires and data dictionaries, and through an iterative process of communication with data contributors, we mapped these elements to common data elements. Definitions, permissible values, and standardized coding were implemented into a single database via SAS and T-SQL. We checked the resulting data for errors and outlying values within and between studies.

We selected environmental factors based on previously being established for their association with CRC risk and on the availability of pooled data in our consortium study.[23–25] Environmental factors included are: anthropometric measurements including body mass index (BMI) per 5 kg/m^2 and height in 10 cm, ever smoked (yes/no), study-specific definitions of regular use of aspirin and nonaspirin nonsteroidal anti-inflammatory drugs (NSAIDs, yes/no), history of type II diabetes (yes/no), and dietary intake. Dietary variables were measured using food frequency questionnaires or diet histories, including alcohol, fruits, vegetables, dietary fiber, red meat, processed meat, dietary and supplemental calcium, dietary and supplemental folate, and total energy intake. We created sex- and study-specific quartiles for all dietary variables except alcohol. We categorized alcohol by grams of alcohol intake per day: nondrinker, 1–28 (moderate drinker), and >28 g/day (heavy drinker). For dichotomous or categorical variables, the lowest category of exposure (or no use) was used as the reference, except for alcohol, where the moderate drinker group was set as the reference, given the observed J-shape association.[24,26]

Polygenic Risk Score Construction

Details on genotyping and quality control have been previously published,[4] and genotyping platforms used are summarized in eTable 1; http://links.lww.com/EDE/C195. Briefly, we excluded genotyped single nucleotide polymorphisms (SNPs) based on call rate (<95–98%), lack of Hardy Weinberg equilibrium ($P < 1 \times 10^{-4}$), inconsistencies between self-reported and genotypic sex, and discordant genotype calls within duplicate samples. We imputed all autosomal SNPs of all studies to the Haplotype Reference Consortium r1.1[27] reference panel via the Michigan Imputation Server[28] and converted them into a binary format for data management and analysis using the R package BinaryDosage.[29] To capture common genetic predisposition, we calculated a genetic risk score combining the estimated effects of 141 previously GWAS-identified SNPs associated with CRC risk.[4,30–32] These SNPs had independent contributions to CRC risk confirmed by linkage disequilibrium and/or conditional association analyses, which accounted for the lead SNP within each region.[4,30–32] The detailed characteristics of these SNPs are provided in eTable 2; http://links.lww.com/EDE/C195. We coded each SNP variable as the expected number of copies of the variant allele. The weights were determined by the marginal log- ORs estimated from prior studies to avoid potential bias. For the known loci identified through GECCO, Colon Cancer Family Registry, or Colorectal Transdisciplinary Study studies, the estimates were adjusted for the winner's curse.[33] The winner's curse phenomenon occurs when the true effect sizes are overestimated due to selection bias from using the same data to identify variants that reach a certain significance threshold.[33,34] We constructed a genetic risk score for each individual by taking the weighted sum of variant alleles over all 141 SNPs, accounting for the strength of CRC association with each SNP. We standardized the genetic risk score as (risk score-mean risk score)/standard deviation (SD), where mean and SD were the mean and SD of the genetic risk score calculated based on all participants.

Additive and Multiplicative Interaction

Consider a binary genotype G and a binary environmental exposure E. Let Y indicate disease status, we can fit a logistic regression with interaction term:

$$logit\{P(Y = 1|G,\ E\} = \beta_0 + \beta_1 G + \beta_2 E + \beta_3 GE, \quad (1)$$

where β_0 is the intercept, β_1 and β_2 are the main effects of G and E, respectively, and β_3 is the multiplicative interaction effect. To measure the interaction on an additive scale, we constructed a 2 × 2 table (Table 1) to facilitate understanding of RERI estimation. The cells in the table represent the OR of

TABLE 1. 2 × 2 table of two exposures (genetic and environmental) and corresponding odds ratios

		Exposure 1 (G)	
		0 (mean)	**1 (mean + 1 SD)**
Exposure 2 (E)	0	1.0 (Ref)	$OR_{01} = e^{\beta_1}$
	1	$OR_{10} = e^{\beta_2}$	$OR_{11} = e^{\beta_1+\beta_2+\beta_3}$

$$\text{logit}\{P(Y = 1 \mid G, E)\} = \beta_0 + \beta_1 G + \beta_2 E + \beta_3 GE \quad (1)$$

Additive Interaction, $\text{RERI} = (OR_{11} - OR_{01}) - (OR_{10} - 1) = \left(e^{\beta_1+\beta_2+\beta_3} - e^{\beta_1}\right) - \left(e^{\beta_2} - 1\right)$ (2)

Multiplicative interaction $= OR_{11}/(OR_{01} * OR_{10}) = e^{\beta_3}$

For continuous genetic exposure, G = 0 is the mean, compared to G = 1 is mean + 1 standard deviation. 0,1 notation is used for simplicity.

disease for the reference group when both G and E are absent ($OR_{00} = 1$), when factor G is present but E is absent (OR_{01}), when factor G is absent but E is present (OR_{10}), and when both factors G and E are present (OR_{11}), respectively. For simplicity we have defined G with binary classifications; in practice, however, in our manuscript G = 0 would represent the mean genetic risk score, and G = 1 represents the mean genetic risk score + one SD. The RERI is then defined as:

$$\text{RERI} = (OR_{11} - OR_{01}) - (OR_{10} - 1) = \left(e^{\beta_1+\beta_2+\beta_3} - e^{\beta_1}\right) - \left(e^{\beta_2} - 1\right) \quad (2)$$

The first part of the equation, $\left(e^{\beta_1+\beta_2+\beta_3} - e^{\beta_1}\right)$, estimates the difference in the OR of developing disease when exposed to environmental exposure compared to when not exposed among individuals who carry variant genotypes. The second part of the equation, $\left(e^{\beta_2} - 1\right)$, estimates the difference in OR when exposed to the environmental factor compared to when not exposed among those who do not carry variant genotypes. Finally, the difference between the two parts estimates excessive OR differences between being exposed and not for those who carry the high-risk genotypes and for those who do not. In other words, RERI quantifies the excess of OR due to jointly being exposed and carrying the high-risk genotype compared to the reference group beyond the summation of the separate excess of ORs of being exposed and carrying the high-risk genotype in the absence of the other risk factor compared to the reference group. Given that the OR approximates the relative risk, the RERI derived from ORs will approximately equal the RERI derived from risk data.[8] We used the delta method for estimating the variance and 95% confidence intervals (CI) of RERI.[35]

The reference level of continuous exposures is of critical importance for the additive interaction RERI, because it is calculated post hoc from a multiplicative logistic regression model. Changing the reference level by adding or subtracting a constant does not change the estimate for multiplicative interaction but does so for additive interaction. This is because additive interaction also involves estimates of main effects, which change if the reference level changes. For continuous variables (e.g., genetic risk score), we used the mean as the reference level so that RERI would reflect the comparison with the mean and the mean generally is more robust than using other values as the reference level (e.g., a minimum value).[36] For dichotomous or categorical variables, no use or lowest quartile was used as the reference regardless of whether the exposure was a risk or protective factor. The exception was alcohol where we set the moderate drinker group as the reference due to the observed J-shape association.[24,26] More details are provided in the eAppendix; http://links.lww.com/EDE/C184.

Statistical Analysis

We presented mean (SD) and counts (percentages) for the genetic risk score and environmental factors for CRC cases and controls separately. We used logistic regression models to calculate OR and 95% CI to assess the associations of the 13 environmental factors and genetic risk scores with CRC risk, whereas for genetic risk score we calculated the OR per 1 SD increase as commonly used for genetic risk score in the literature. In our models, quartile variables (Q2, Q3, and Q4) were compared to Q1 as the reference group, treating each quartile exposure variable as a categorical predictor. We also used the ordinal quartile variable (1, 2, 3, and 4) to calculate a linear trend test. Our primary complete data analysis models were adjusted for age at reference time, sex, study, the first three principal components of genetic ancestry, and total energy for dietary variables only. In further models, we adjusted for all other 12 environmental factors of interest, where we used study- and sex-specific mean imputation for missing values (missing rates <5%) in adjusted covariates to maximize our sample size. In our previous work, we have found consistent results utilizing mean versus multiple imputation; therefore, we used mean imputation here for simplicity.[37] We also assessed the correlation between genetic risk score and environmental factors by using a linear regression model among controls for the association of environmental factors with genetic risk score, adjusted for age at reference time, sex, study, the first three principal components of genetic ancestry, and total energy for dietary variables only. For BMI and height, we stratified the analyses by sex, given that BMI and height vary by sex. We additionally conducted a sensitivity analysis restricted to colon-only cancers (26,102 colon cancer cases and 48,116 controls). Forest plots of additive (RERI) and multiplicative interaction estimates and 95% CI were plotted. To account for multiple testing of 11 exposure variables and two sex-stratified exposure variables, the Holm-Bonferroni method for *P*-value correction was used with 0.05 as the overall significance threshold for additive and multiplicative interaction separately (presented in eTable 4; http://links.lww.com/EDE/C184). All analyses were performed using R, version 4.1.3 (R Foundation for Statistical Computing, Vienna, Austria) software.

RESULTS

Our study included a total of 97,918 participants with 45,247 cases and 52,671 controls (eTable 1; http://links.lww.com/EDE/C195). Heavy or no alcohol, smoking, diabetes, high BMI for both sexes, taller height among females, and high processed meat and red meat consumption were associated with higher risk of CRC (Table 2). Use of aspirin/NSAIDs,

TABLE 2. Comparison by Colorectal Cancer Case–control Status and Association (Odds Ratio, 95% Confidence Interval) of Genetic Risk Score, Environmental Factors, and Risk of Colorectal Cancer

	Overall	Cases	Controls	OR (95% CI)[a]
Genetic risk score, mean (std)	9.3 (0.51)	9.4 (0.50)	9.2 (0.50)	1.5 (1.5, 1.6) per 1 SD
Risk factors				
Alcohol nondrinker (<1 g/day), N (%)	29,293 (40)	13,754 (43)	15,539 (37)	1.2 (1.1, 1.2)
Alcohol 1–28 g/day, N (%)	35,635 (48)	13,979 (44)	21,656 (51)	1
Alcohol >28 g/day, N (%)	9162 (12)	4141 (13)	5021 (12)	1.4 (1.3, 1.5)
Ever smoked, no, N (%)	36,947 (47)	14,975 (44)	21,972 (50)	1
Ever smoked, yes, N (%)	41,144 (53)	18,781 (56)	22,363 (50)	1.2 (1.2, 1.3)
History of type II diabetes, no, N (%)	67,524 (91)	27,957 (89)	39,567 (92)	1
History of type II diabetes, yes, N (%)	6854 (9)	3572 (11)	3282 (8)	1.3 (1.3, 1.4)
BMI, female, mean (std)[b]	27 kg/m^2 (5.2)	27 kg/m^2 (5.4)	27 kg/m^2 (5.1)	1.1 (1.1, 1.1) per 5 kg/m^2
BMI, male, mean (std)[b]	27 kg/m^2 (4.2)	28 kg/m^2 (4.3)	27 kg/m^2 (4.1)	1.2 (1.2, 1.2) per 5 kg/m^2
Height, female, mean (std)[b]	1.6 m (0.1)	1.6 m (0.1)	1.6 m (0.1)	1.1 (1.0, 1.1) per 10 cm
Height, male, mean (std)[b]	1.8 m (0.1)	1.8 m (0.1)	1.8 m (0.1)	1.0 (1.0, 1.0) per 10 cm
Red meat, Quartile 1 (lowest), N (%)[c]	18,665 (24)	7482 (23)	11,183 (26)	1
Red meat, Quartile 2, N (%)	22,204 (29)	9243 (28)	12,961 (30)	1.1 (1.1, 1.2)
Red meat, Quartile 3, N (%)	19,869 (26)	8517 (26)	11,352 (26)	1.2 (1.2, 1.3)
Red meat, Quartile 4 (highest), N (%)	15,580 (20)	7645 (23)	7935 (18)	1.3 (1.3, 1.4)
P_{trend}				6.65E-35
Processed meat, Quartile 1 (lowest), N (%)[c]	12,761 (18)	5084 (17)	7677 (19)	1
Processed meat, Quartile 2, N (%)	24,160 (35)	9947 (33)	14,213 (35)	1.1 (1.0, 1.1)
Processed meat, Quartile 3, N (%)	22,993 (33)	9941 (33)	13,052 (32)	1.2 (1.2, 1.3)
Processed meat, Quartile 4 (highest), N (%)	9998 (14)	4762 (16)	5236 (13)	1.2 (1.1, 1.30)
P_{trend}				5.73E-19
Protective factors				
Regular aspirin/NSAIDs use at reference time, No, N (%)	45,396 (62)	20,372 (66)	25,024 (60)	1
Regular aspirin/NSAIDs use, Yes, N (%)	27,290 (38)	10,450 (34)	16,840 (40)	0.76 (0.73, 0.79)
Vegetable, Quartile 1 (lowest), N (%)[c]	17,528 (23)	7393 (23)	10,135 (23)	1
Vegetable, Quartile 2, N (%)	23,722 (31)	11,121 (34)	12,601 (29)	0.93 (0.89, 0.97)
Vegetable, Quartile 3, N (%)	18,952 (25)	7944 (24)	11,008 (25)	0.84 (0.81, 0.88)
Vegetable, Quartile 4, N (%)	15,943 (21)	6401 (19)	9542 (22)	0.83 (0.79, 0.87)
P_{trend}				2.21E-18
Fruit, Quartile 1 (lowest), N (%)[c]	19,272 (25)	8762 (27)	10,510 (24)	1
Fruit, Quartile 2, N (%)	23,449 (31)	10,626 (32)	12,823 (30)	0.85 (0.82, 0.89)
Fruit, Quartile 3, N (%)	17,733 (23)	7322 (22)	10,411 (24)	0.77 (0.73, 0.80)
Fruit, Quartile 4 (highest), N (%)	15,548 (20)	6071 (19)	9477 (22)	0.74 (0.70, 0.77)
P_{trend}				1.30E-43
Fiber (g/day), Quartile 1 (lowest), N (%)[c]	12,231 (25)	6160 (26)	6071 (23)	1
Fiber, Quartile 2, N (%)	12,607 (25)	6001 (26)	6606 (25)	0.84 (0.80, 0.89)
Fiber, Quartile 3, N (%)	12,404 (25)	5637 (24)	6767 (26)	0.73 (0.69, 0.77)
Fiber, Quartile 4 (highest), N (%)	12,493 (25)	5583 (24)	6910 (26)	0.65 (0.61, 0.69)
P_{trend}				2.62E-47
Total folate (mcg/day), Quartile 1 (lowest), N (%)[c]	14,132 (24)	6961 (25)	7171 (23)	1
Total folate, Quartile 2, N (%)	15,797 (27)	7556 (27)	8241 (27)	0.86 (0.82, 0.90)
Total folate, Quartile 3, N (%)	13,566 (23)	6443 (23)	7123 (23)	0.83 (0.79, 0.87)
Total folate, Quartile 4 (highest), N (%)	14,969 (26)	6695 (24)	8274 (27)	0.75 (0.71, 0.79)
P_{trend}				1.44E-24
Dietary folate (mcg/day), Quartile 1 (lowest), N (%)[c]	12,488 (24)	5946 (25)	6542 (24)	1
Dietary folate, Quartile 2, N (%)	12,894 (25)	5879 (25)	7015 (25)	0.88 (0.83, 0.92)
Dietary folate, Quartile 3, N (%)	13,157 (25)	6010 (25)	7147 (26)	0.84 (0.80, 0.89)
Dietary folate, Quartile 4 (highest), N (%)	13,110 (25)	5995 (25)	7115 (26)	0.79 (0.74, 0.84)
P_{trend}				8.27E-13
Total calcium (mg/day), Quartile 1 (lowest), N (%[c]	14,466 (19)	7398 (22)	7068 (16)	1
Total calcium, Quartile 2, N (%)	29,355 (38)	11,534 (34)	17,821 (41)	0.87 (0.83, 0.91)

(Continued)

TABLE 2. Continued

	Overall	Cases	Controls	OR (95% CI)[a]
Total calcium, Quartile 3, N (%)	18,873 (24)	8216 (24)	10,657 (24)	0.77 (0.73, 0.81)
Total calcium, Quartile 4, N (%)	14,861 (19)	6583 (20)	8278 (19)	0.67 (0.64, 0.71)
P_{trend}				1.11E-55
Dietary calcium (mg/day), Quartile 1 (lowest), N (%)[c]	13,860 (24)	6993 (25)	6867 (23)	1
Dietary calcium, Quartile 2, N (%)	14,616 (26)	7198 (26)	7418 (25)	0.91 (0.87, 0.96)
Dietary calcium, Quartile 3, N (%)	14,200 (25)	6573 (24)	7627 (26)	0.77 (0.73, 0.81)
Dietary calcium, Quartile 4, N (%)	14,514 (25)	6691 (24)	7823 (26)	0.72 (0.68, 0.76)
P_{trend}				3.28E-35

[a]Logistic regression adjusted for age, sex, study, the first three principal components of ancestry, and total energy for dietary variables only.
[b]BMI data was complete for 36,415 cases and 48,439 controls. Height data was complete for 39,798 cases and 50,340 controls.
[c]Quartiles are study and sex-specific.

higher intakes of dietary and total calcium, fiber, dietary/total folate, vegetable, and fruit were associated with lower risk of CRC. Genetic risk score was associated with a 1.5-fold higher risk of CRC per SD increase in genetic risk score (95% CI = 1.5, 1.6). There was no evidence of the correlation between genetic risk score and any of the environmental risk factors (eTable 3; http://links.lww.com/EDE/C184).

Interaction of Environmental Risk Factors and Genetic Risk Score with CRC Risk

Across all environmental risk factors assessed, there were no large magnitude multiplicative interactions with genetic risk on CRC risk (all OR estimates >0.95 and <1.05, Figure A, right panel). In contrast, we observed large-magnitude additive interaction effects for several environmental factors: heavy alcohol consumption, ever-smoking status, BMI, and red meat (Figure A, left panel, eTable 4; http://links.lww.com/EDE/C184). Specifically, the RERI (95% CI) for heavy alcohol consumption was 0.24 (0.13, 0.36); ever smoking, 0.11 (0.05, 0.16); for BMI among females 0.09 (0.05, 0.13) per 5 kg/m^2 increase, among males 0.10 (0.05, 0.14) per 5 kg/m^2 increase; and red meat intake, RERI Q4 (highest) versus Q1 (lowest) 0.18 (0.09, 0.27). The joint effect of each of these risk factors and the high genetic risk score on CRC risk was higher than the expected sum of individual effects. We additionally observed additive interaction effects for alcohol nondrinkers and processed meat intake, where processed meat intake RERI had a relatively large estimate (RERI [95% CI] processed meat Q4 [highest] vs Q1 [lowest] = 0.15 [0.04, 0.26], Figure A, left panel, eTable 4; http://links.lww.com/EDE/C184). We conducted a sensitivity analysis using further adjusted models and the results did not materially change (eTable 5; http://links.lww.com/EDE/C184). In the sensitivity analyses restricted to colon cancer only, results were similar and, in some cases, stronger than for CRC cases overall (eTable 6; http://links.lww.com/EDE/C184).

Interaction of Environmental Protective Factors and Genetic Risk Score With CRC Risk

Across all environmental protective factors assessed, there were no large magnitude multiplicative interactions with genetic risk score on CRC risk (all OR estimates >0.95 and <1.05, Figure B, right panel). We observed large magnitude additive interaction for several protective factors: aspirin/NSAIDs, fruit, fiber, and total calcium (Figure B, left panel, eTable 4; http://links.lww.com/EDE/C184). The RERI (95% CI) for these protective factors were use of aspirin or NSAIDs −0.16 (−0.2, −0.11); fruit Q4 (highest) versus Q1 (lowest) −0.12 (−0.18, −0.05); fiber Q4 (highest) versus Q1 (lowest) −0.16 (−0.23, −0.09); and total calcium Q4 (highest) versus Q1 (lowest) −0.11 (−0.18, −0.05), respectively. The beneficial effects of these environmental factors may be more protective against CRC risk among those with high genetic risk score compared to those with low-risk score on the additive scale. We additionally observed additive interaction for vegetable, total folate, dietary folate, and dietary calcium intake, as well as Q3 versus Q1 for total fiber and total calcium, where all RERI estimates ranged from −0.09 to −0.11 (Figure B, left panel, eTable 4; http://links.lww.com/EDE/C184). We conducted a sensitivity analysis using further adjusted models and the results did not materially change (eTable 5; http://links.lww.com/EDE/C184). In the colon-only sensitivity analyses, results were also similar and, in some cases, stronger than for CRC cases overall (eTable 6; http://links.lww.com/EDE/C184).

DISCUSSION

In this large study from international CRC consortia where we aimed to assess gene–environment interactions for subgroup identification, we observed additive interactions between several risk and protective environmental factors and elevated genetic risk score on CRC risk. These results are consistent with the hypothesis that, for individuals with high genetic susceptibility, either heavy drinking, ever smoking, high BMI, or high red meat intake confer excessive CRC risk greater than that for individuals with average genetic susceptibility. In other words, the excess relative risk associated with the genetic risk score and the respective environmental factor together is greater than the sum of the excessive relative risk of the genetic risk score and the respective environmental factor separately.

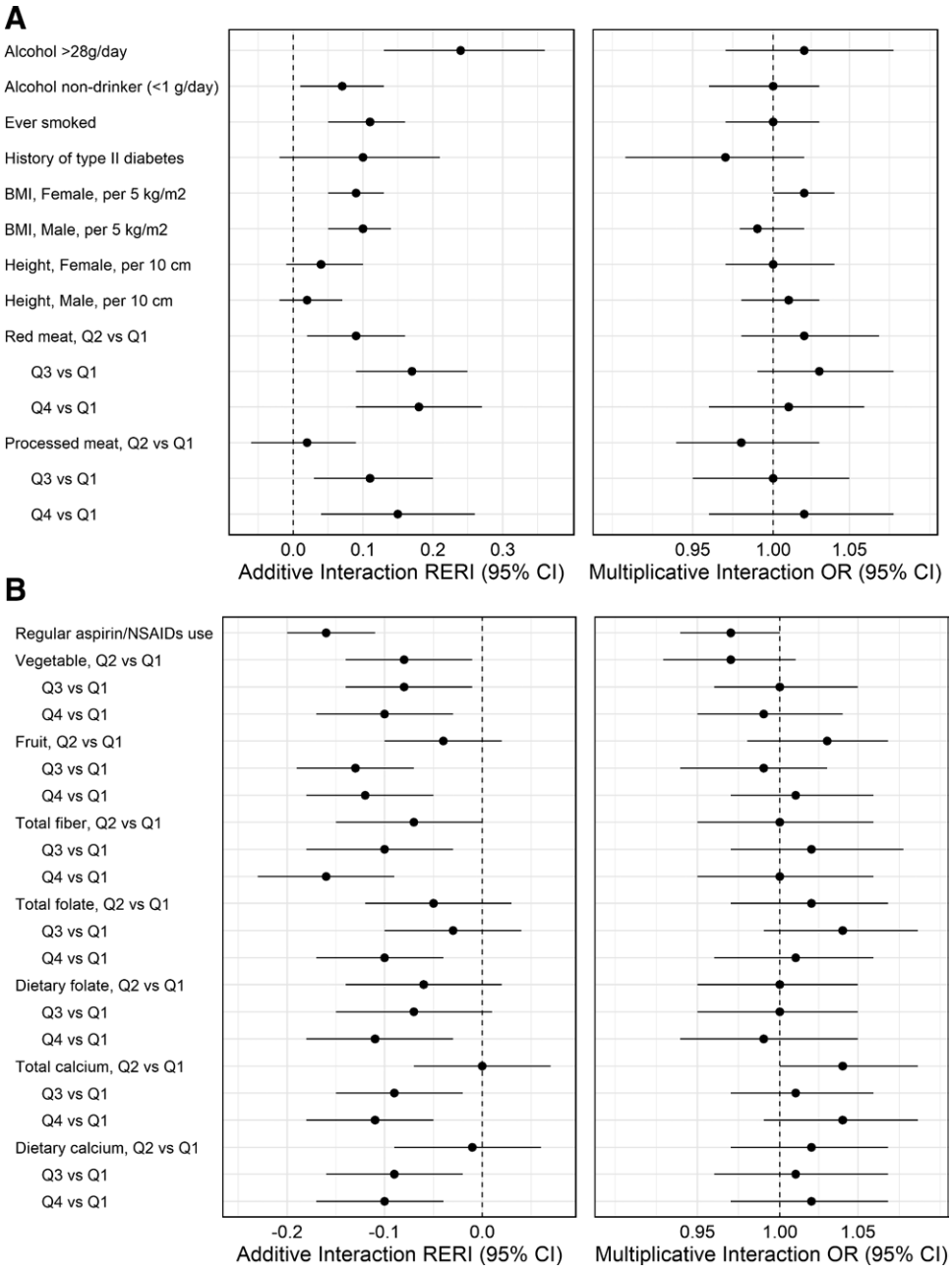

FIGURE. Estimate for additive and multiplicative interactions between genetic risk score and (a) risk factors for colorectal cancer, (b) protective factors for colorectal cancer. Logistic regression adjusted for age, sex, study, first three principal components of ancestry, and total energy for dietary variables only.

Conversely, we estimate that those with high genetic susceptibility experience greater benefits of reducing CRC risk with aspirin or NSAID use or higher intake of fruit, fiber, or calcium than those with average genetic susceptibility. While environmental factors show a consistent direction of association regardless of genetic susceptibility, these results can be useful in identifying those at high genetic risk who may benefit more from environmental factor intervention than the general population, as well as inform personalized healthcare decision-making.

Our findings are consistent with a previous study on additive interaction in CRC, which found that healthy lifestyle score and genetic risk score had an additive interaction where those with high genetic risk score and unhealthy healthy lifestyle score had a relative excess risk due to the interaction of 0.58 (95% CI = 0.06, 1.10) compared to the sum of risk from either high genetic risk score or unhealthy healthy lifestyle score.[20] Other articles examining multiplicative interaction with CRC-related genetic risk score found, consistent with our findings, no interaction with smoking,[12] NSAIDs,[13] fine

particulate matter,[14] aspirin,[16] red and processed meat intake,[17] alcohol consumption,[18] physical activity,[19] or several environmental factors including height, BMI, smoking, alcohol, vegetable, fruit, processed meat, fiber, and others.[15] It is worth noting that for the GxE interaction studies focusing on single SNPs, there is little or no difference between additive and multiplicative interaction due to weak SNP effect size, as commonly observed.[38] However, when genetic risk score is used to capture overall genetic susceptibility, the effect size of the genetic risk score is sizable, and the difference between multiplicative interaction and RERI may be substantial, as shown in our study, where we observed additive interactions but no multiplicative interactions for many environmental risk factors. This observation indicates the importance of assessing interaction on both additive and multiplicative scales, where the additive results, proportional to excess absolute risk, may be more useful to public health intervention.

Our study improves upon current findings by suggesting which environmental factors specifically have evidence of GxE interaction, as those factors may be better targets through which to intervene on the population level. Many environmental factor recommendations apply to the general population, regardless of genetic risk: less heavy alcohol consumption, no smoking, lower adiposity, less red meat, and more fiber, fruit, and calcium. However, having knowledge of which subgroups may benefit more from encouraging better lifestyle factors may help inform personalized intervention and medical decision-making. For example, NSAIDs are inversely associated with CRC risk; however, it also carries risk from potential side effects. Identifying those who may benefit more from NSAIDs while accounting for side effects may help inform an individual's risk and benefit decision-making about the use of such drugs.

We conducted a sensitivity analysis using further adjusted models to verify that our additive interaction results were unlikely to be biased by confounding of environmental factors. Our results were broadly consistent between minimally adjusted and further adjusted models, suggesting that environmental confounding had a limited impact on our interaction findings. Under the assumptions of gene–environment independence, a rare outcome, and no interaction between unmeasured confounders and genetic risk score, the multiplicative interaction term will be valid even in the presence of an unmeasured or incompletely adjusted for environmental confounder.[39] In our study, there was little evidence of dependence between genetic risk score and environmental factors (eTable 3; http://links.lww.com/EDE/C184). However, the main effects of E may be biased by potential environmental confounding.[39] Given that we derive additive interaction post hoc from a multiplicative logistic regression model, it uses both the main effects and interaction term in calculation of RERI. Thus, additive interaction is more likely to be affected by environmental confounding compared to multiplicative interaction, where the multiplicative interaction term may still be valid. However, given similar results with further adjustment of all other environmental factors, confounding does not appear to play a substantial role.

RERI is on the OR scale and may be more generalizable to other populations than absolute risk. This is because absolute risk entails baseline absolute risk, and accurately estimating it can be challenging due to population heterogeneity. On the other hand, as RERI is calculated based on the logistic regression model, which is often adjusted for additional covariates (e.g., age and sex) to account for confounding, interpreting the RERI magnitude in this context requires caution. The excess risk or protection effect implied by RERI may vary across covariate values, even though the RERI itself remains constant. Nonetheless, the direction of RERI and its implications in the public health relevance should remain consistent.[8] It is arguable whether a logistic regression model is suitable for assessing additive interaction due to potential model misspecification.[7] However, logistic regression is a well-tested approach for analyzing retrospectively collected case–control data. While caution should be taken, as is the case with any modeling, RERI-derived post hoc from logistic regression may be meaningful, particularly in the identification of high-risk subgroups for public health impact.

Our study has several strengths. We investigated gene–environment interactions on both the multiplicative and additive scales. The inclusion of additive interaction improves our ability to identify subgroups that may benefit the most from public health intervention.[8] We also utilized a comprehensive genetic risk score of CRC risk based on known loci from prior studies, including 141 GWAS-identified variants, using a large sample size. Our study also has limitations, the chief of which is that our study only included individuals of European ancestry, limiting the generalizability of our findings to other racial/ethnic groups. Further studies of GxE in diverse populations are needed to equitably identify relevant subgroups that may benefit the most from additional intervention. Of note, the effect size of the genetic risk score is reflective of the combined influence of multiple genetic variants and is expressed in terms of SDs, where the interaction effects, whether additive or multiplicative, would be reported in terms of SDs of the genetic risk score, providing a relative measure of the influence of genetic profile on disease risk. Genetic risk score distributions may vary across populations due to differential linkage disequilibrium and allele frequencies. As a result, the effect size of genetic risk score per SD may vary across populations. Although our findings in the European ancestry population likely hold qualitatively, it is important to assess the multiplicative and additive interactions in other racial and ethnic groups. Additionally, environmental factors were measured at only one single time point and may not capture their effects on CRC risk adequately. However, we used a standardized protocol to harmonize environmental factors and our previous work based on this large pooled study that included both retrospective and prospective studies showed

consistent associations of environmental factors with CRC risk,[24,37,40] supporting the robustness of the results and pooling of these studies. Some GWAS summary statistics used to create the genetic risk score overlap with GECCO studies,[4] which may cause inflation of the genetic risk score main effect. This is an example of the "winner's curse" phenomenon, where the true effect sizes are overestimated due to selection bias from using the same data to identify variants that reach a certain significance threshold.[33,34] While the effect size estimates for constructing the genetic risk score were adjusted for winner's curse[33] and the main effect of genetic risk score could still be overestimated because the same data was used, the (multiplicative) GxE interaction may not be. Therefore, while it is possible that the magnitude of additive interaction is overestimated, it is unlikely that the conclusion is wrong qualitatively. As previously mentioned, it is important to note that when assessing interaction, if both exposures of interest have an effect on the outcome interaction must be present on at least one scale or both.[8] Given that we selected genetic risk score and environmental factors based on their well-established association with CRC, our results of exposures showing interaction effects are expected. However, the overall lack of a multiplicative interaction is intriguing. This could be because of lower power on the multiplicative scale than on the additive scale.[41] Nevertheless, we note that the estimates of multiplicative interaction were close to the null, and the 95% CIs well covered the null. In contrast, the 95% CIs for additive interaction were generally away from the null.

In conclusion, we identified additive interaction between multiple risk and protective environmental factors and genetic risk scores on CRC risk in individuals of European descent. These findings may help inform which population subgroups may gain additional benefit from environmental factor intervention utilizing genetic information in practice. Future work is needed in risk prediction and intervention of modifiable risk factors among genetically susceptible high-risk subgroups.

ACKNOWLEDGMENTS

ASTERISK: We are very grateful to Dr. Bruno Buecher, without whom this project would not have existed. We also thank all those who agreed to participate in this study, including the patients and the healthy control persons, as well as all the physicians, technicians, and students.

CCFR: The Colon CFR graciously thanks the generous contributions of their study participants, dedication of study staff, and the financial support from the US National Cancer Institute, without which this important registry would not exist.

CLUE II: We thank the participants of Clue II and appreciate the continued efforts of the staff at the Johns Hopkins George W. Comstock Center for Public Health Research and Prevention in the conduct of the Clue II Cohort Study. Cancer data was provided by the Maryland Cancer Registry, Center for Cancer Prevention and Control, Maryland Department of Health, with funding from the State of Maryland and the Maryland Cigarette Restitution Fund. The collection and availability of cancer registry data is also supported by the Cooperative Agreement NU58DP006333, funded by the Centers for Disease Control and Prevention. Its contents are solely the responsibility of the authors and do not necessarily represent the official views of the Centers for Disease Control and Prevention or the Department of Health and Human Services.

CORSA: We kindly thank all individuals who agreed to participate in the CORSA study. Furthermore, we thank all cooperating physicians and students and the Biobank Graz of the Medical University of Graz.

CPS-II: The authors express sincere appreciation to all Cancer Prevention Study-II participants, and to each member of the study and biospecimen management group. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries and cancer registries supported by the National Cancer Institute's Surveillance Epidemiology and End Results Program. The authors assume full responsibility for all analyses and interpretations of results. The views expressed here are those of the authors and do not necessarily represent the American Cancer Society or the American Cancer Society – Cancer Action Network.

Czech Republic CCS: We are thankful to all clinicians in major hospitals in the Czech Republic, without whom the study would not be practicable. We are also sincerely grateful to all patients participating in this study.

DACHS: We thank all participants and cooperating clinicians, and everyone who provided excellent technical assistance.

EDRN: We acknowledge all contributors to the development of the resource at the University of Pittsburgh School of Medicine, Department of Gastroenterology, Department of Pathology, Hepatology and Nutrition, and Biomedical Informatics.

EPIC: Where authors are identified as personnel of the International Agency for Research on Cancer/World Health Organization, the authors alone are responsible for the views expressed in this article and they do not necessarily represent the decisions, policy, or views of the International Agency for Research on Cancer/World Health Organization.

EPICOLON: We are sincerely grateful to all patients participating in this study who were recruited as part of the EPICOLON project. We acknowledge the Spanish National DNA Bank, Biobank of Hospital Clínic–IDIBAPS, and Biobanco Vasco for the availability of the samples. The work was carried out (in part) at the Esther Koplowitz Centre, Barcelona.

Harvard cohorts: The study protocol was approved by the institutional review boards of the Brigham and Women's Hospital and Harvard T.H. Chan School of Public Health, and those of participating registries as required. We acknowledge Channing Division of Network Medicine, Department

of Medicine, Brigham and Women's Hospital as home of the NHS. The authors would like to acknowledge the contribution to this study from central cancer registries supported through the Centers for Disease Control and Prevention's National Program of Cancer Registries (NPCR) and/or the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) Program. Central registries may also be supported by state agencies, universities, and cancer centers. Participating central cancer registries include the following: Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Hawaii, Idaho, Indiana, Iowa, Kentucky, Louisiana, Massachusetts, Maine, Maryland, Michigan, Mississippi, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Puerto Rico, Rhode Island, Seattle SEER Registry, South Carolina, Tennessee, Texas, Utah, Virginia, West Virginia, Wyoming. The authors assume full responsibility for analyses and interpretation of these data.

Kentucky: We would like to acknowledge the staff at the Kentucky Cancer Registry.

LCCS: We acknowledge the contributions of Jennifer Barrett, Robin Waxman, Gillian Smith, and Emma Northwood in conducting this study.

NCCCS I & II: We would like to thank the study participants and the NC Colorectal Cancer Study staff.

PLCO: The authors thank the PLCO Cancer Screening Trial screening center investigators and the staff from Information Management Services Inc. and Westat Inc. Most importantly, we thank the study participants for their contributions that made this study possible.

SEARCH: We thank the SEARCH team.

SELECT: We thank the research and clinical staff at the sites that participated in the SELECT study, without whom the trial would not have been successful. We are also grateful to the 35,533 dedicated men who participated in SELECT.

WHI: The authors thank the WHI investigators and staff for their dedication, and the study participants for making the program possible. A full listing of WHI investigators can be found at: https://www.whi.org/doc/WHI-Investigator-Short-List.pdf

REFERENCES

1. Sung H, Ferlay J, Siegel RL, et al. Global Cancer Statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin*. 2021;71:209–249.
2. Lichtenstein P, Holm NV, Verkasalo PK, et al. Environmental and heritable factors in the causation of cancer — analyses of cohorts of twins from Sweden, Denmark, and Finland. *N Engl J Med*. 2000;343:78–85.
3. Mucci LA, Hjelmborg JB, Harris JR, et al; Nordic Twin Study of Cancer (NorTwinCan) Collaboration. Familial risk and heritability of cancer among twins in Nordic Countries. *JAMA*. 2016;315:68–76.
4. Huyghe JR, Bien SA, Harrison TA, et al. Discovery of common and rare genetic risk variants for colorectal cancer. *Nat Genet*. 2019;51:76–87.
5. Fernandez-Rozadilla C, Timofeeva M, Chen Z, et al. Deciphering colorectal cancer genetics through multi-omic analysis of 100,204 cases and 154,587 controls of European and East Asian ancestries. *Nat Genet*. 2023;55:89–99.
6. McAllister K, Mechanic LE, Amos C, et al. Current challenges and new opportunities for gene–environment interaction studies of complex diseases. *Am J Epidemiol*. 2017;186:753–761.
7. Skrondal A. Interaction as departure from additivity in case–control studies: a cautionary note. *Am J Epidemiol*. 2003;158:251–258.
8. Lash TL, VanderWeele TJ, Haneuse S, Rothman KJ. *Modern Epidemiology*. 4th ed. LWW; 2021.
9. Greenland S, Lash TL, Rothman KJ. "Concepts of interaction," chapter 5. In: Rothman KJ, Greenland S, Lash TL, eds. *Modern Epidemiology*. 3rd ed. 2008. Philadelphia, PA: Lippincott Williams and Wilkins.
10. VanderWeele TJ, Knol MJ. A tutorial on interaction. *Epidemiol Methods*. 2014;3:33–72.
11. Knol MJ, VanderWeele TJ. Recommendations for presenting analyses of effect modification and interaction. *Int J Epidemiol*. 2012;41:514–520.
12. Chen X, Jansen L, Guo F, Hoffmeister M, Chang-Claude J, Brenner H. Smoking, genetic predisposition, and colorectal cancer risk. *Clin Transl Gastroenterol*. 2021;12:e00317.
13. Chen X, Guo F, Hoffmeister M, Chang-Claude J, Brenner H. Nonsteroidal anti-inflammatory drugs, polygenic risk score and colorectal cancer risk. *Aliment Pharmacol Ther*. 2021;54:167–175.
14. Chu H, Xin J, Yuan Q, et al. A prospective study of the associations among fine particulate matter, genetic variants, and the risk of colorectal cancer. *Environ Int*. 2021;147:106309.
15. Yang T, Li X, Farrington SM, et al. A systematic analysis of interactions between environmental risk factors and genetic variation in susceptibility to colorectal cancer. *Cancer Epidemiol Biomarkers Prev*. 2020;29:1145–1153.
16. Bakshi A, Cao Y, Orchard SG, et al. Aspirin and the risk of colorectal cancer according to genetic susceptibility among older individuals. *Cancer Prev Res (Phila)*. 2022;15:447–454.
17. Chen X, Hoffmeister M, Brenner H. Red and processed meat intake, polygenic risk score, and colorectal cancer risk. *Nutrients*. 2022;14:1077.
18. Chen X, Li H, Guo F, Hoffmeister M, Brenner H. Alcohol consumption, polygenic risk score, and early- and late-onset colorectal cancer risk. *EClinicalMedicine*. 2022;49:101460.
19. Chen X, Guo F, Chang-Claude J, Hoffmeister M, Brenner H. Physical activity, polygenic risk score, and colorectal cancer risk. *Cancer Med*. 2023;12:4655–4666.
20. Choi J, Jia G, Wen W, Shu XO, Zheng W. Healthy lifestyles, genetic modifiers, and colorectal cancer risk: a prospective cohort study in the UK Biobank. *Am J Clin Nutr*. 2021;113:810–820.
21. Schmit SL, Edlund CK, Schumacher FR, et al. Novel common genetic susceptibility loci for colorectal cancer. *J Natl Cancer Inst*. 2019;111:146–157.
22. Schumacher FR, Schmit SL, Jiao S, et al. Genome-wide association study of colorectal cancer identifies six new susceptibility loci. *Nat Commun*. 2015;6:7138.
23. Hutter CM, Chang-Claude J, Slattery ML, et al. Characterization of gene–environment interactions for colorectal cancer susceptibility loci. *Cancer Res*. 2012;72:2036–2044.
24. McNabb S, Harrison TA, Albanes D, et al. Meta-analysis of 16 studies of the association of alcohol with colorectal cancer. *Int J Cancer*. 2020;146:861–873.
25. Xia Z, Su Y, Petersen P, et al. Functional informed genome-wide interaction analysis of body mass index, diabetes and colorectal cancer risk. *Cancer Med*. 2020;9:3563–3573.
26. Gong J, Hutter CM, Newcomb PA, et al; CCFR and GECCO. Genome-wide interaction analyses between genetic variants and alcohol consumption and smoking for risk of colorectal cancer. *PLoS Genet*. 2016;12:e1006296.
27. McCarthy S, Das S, Kretzschmar W, et al; Haplotype Reference Consortium. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48:1279–1283.
28. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016;48:1284–1287.
29. Morrison J. Binary Dosage: Creates, Merges, and Reads Binary Dosage Files. R package version 1.0.0. 2020. Available at: https://CRAN.R-project.org/package=BinaryDosage. Accessed 19 November 2022.
30. Archambault AN, Jeon J, Lin Y, et al. Risk stratification for early-onset colorectal cancer using a combination of genetic and environmental

risk scores: an international multi-center study. *J Natl Cancer Inst*. 2022;114:528–539.
31. Law PJ, Timofeeva M, Fernandez-Rozadilla C, et al; PRACTICAL consortium. Association analyses identify 31 new risk loci for colorectal cancer susceptibility. *Nat Commun*. 2019;10:2154.
32. Lu Y, Kweon SS, Tanikawa C, et al. Large-scale genome-wide association study of east Asians identifies loci associated with risk for colorectal cancer. *Gastroenterology*. 2019;156:1455–1466.
33. Zhong H, Prentice RL. Bias-reduced estimators and confidence intervals for odds ratios in genome-wide association studies. *Biostatistics (Oxford, England)*. 2008;9:621–634.
34. Capen EC, Clapp RV, Campbell WM. Competitive bidding in high-risk situations. *J Petrol Technol*. 1971;23:641–653.
35. Hosmer DW, Lemeshow S. Confidence interval estimation of interaction. *Epidemiology (Cambridge, Mass.)*. 1992;3:452–456.
36. Knol MJ, VanderWeele TJ, Groenwold RHH, Klungel OH, Rovers MM, Grobbee DE. Estimating measures of interaction on an additive scale for preventive exposures. *Eur J Epidemiol*. 2011;26:433–438.
37. Wang X, O'Connell K, Jeon J, et al. Combined effect of modifiable and non-modifiable risk factors for colorectal cancer risk in a pooled analysis of 11 population-based studies. *BMJ Open Gastroenterol*. 2019;6:e000339.
38. Gauderman WJ, Mukherjee B, Aschard H, et al. Update on the state of the science for analytical methods for gene–environment interactions. *Am J Epidemiol*. 2017;186:762–770.
39. VanderWeele TJ, Ko YA, Mukherjee B. Environmental confounding in gene–environment interaction studies. *Am J Epidemiol*. 2013;178:144–152.
40. Jeon J, Du M, Schoen RE, et al; Colorectal Transdisciplinary Study and Genetics and Epidemiology of Colorectal Cancer Consortium. Determining risk of colorectal cancer and starting age of screening based on lifestyle, environmental, and genetic factors. *Gastroenterology*. 2018;154:2152–2164.e19.
41. Poole C, Shrier I, VanderWeele TJ. Is the risk difference really a more heterogeneous measure? *Epidemiology (Cambridge, Mass.)*. 2015;26:714–718.